



결측 환경에서의 기상 시계열 예측을 위한 중요도 적응형 손실 함수 설계

장영준¹⁾ · 조승빈²⁾ · 김지홍^{3)*}

Importance-Adaptive Loss Function Design for Meteorological Time-Series Forecasting under Missing Data Conditions

Young-Jun Jang¹⁾ · Seung-Bin Cho²⁾ · Jeehong Kim^{3)*}

Received 10 October 2025 Revised 15 February 2026 Accepted 20 March 2026 Published online 25 March 2026

ABSTRACT This study proposes a novel approach that interprets missing values in meteorological observations as changes in data reliability rather than as simple errors. It integrates them into the learning process of time-series forecasting models. Real-world meteorological data can experience both random and long-term missing structural values, owing to factors such as communication and sensor failures. This study analyzed data from the Jeonju and Gochang observatories to identify temporal patterns and characteristics of missing values and to quantify changes in data reliability as a function of missingness rates and sequence length. Based on this analysis, we designed a dynamic weighted mean squared error loss function, *dyn_wmse2*, that combines missingness-mask-based reliability weights with SHAP-based feature importance. The proposed loss function showed the highest prediction stability and lower performance degradation than the existing MSE, MAE, Huber, and static WMSE loss functions, even in environments with increasing missingness rates. The performance improvement was particularly noticeable in block-missing environments, suggesting their potential for practical use in real-world observation systems. This study empirically demonstrated that designing a loss function and improving the learning structure, rather than focusing on handling missingness, is an effective way to improve the reliability of time-series forecasting.

Key words Solar radiation forecasting(일사량 예측), Missing data(결측), Dynamic weighted loss function(동적 손실함수), SHAP(SHAP), Time-series forecasting(시계열 예측)

Nomenclature

$X(i)$: input time-window matrix

$M_{t,f}$: missing-value mask

1) Researcher, Bumhan Unisolutions Inc.

2) Senior Researcher, SSCompany Inc.

3) Professor, Department of Renewable Energy Engineering, Jeonju
Vision College

*Corresponding author: goodjeehong@gmail.com

Tel:

$W_{t,f}$: weight map derived from missingness

$r(i)$: reliability coefficient based on missingness

η_f : SHAP-based importance weight

η_f^{reg} : regularized SHAP importance

$\hat{W}_{t,f}$: combined weight

$w(i)$: sample-level aggregated reliability weight

$\hat{w}(i)$: dynamic sample weight

$\mathcal{L}_{dyn-wmse2}$: proposed dynamic weighted MSE loss

- $A_{i,f}(i)$: grad×input attribution score
- $I(i)$: scalar attribution
- $\hat{w}_{grad}(i)$: dynamic weight using grad×input
- $\mathcal{L}_{dym-wmse(grad)}$: comparison dynamic weighted loss
- L : lookback window length
- F : number of input numerical features
- α, β, λ : regularization, control coefficients

Subscript

- ANN : Artificial Neural Network
- LSTM : Long Short-Term Memory
- SHAP : SHapley Additive exPlanations
- MSE : Mean Squared Error
- MAE : Mean Absolute Error
- Huber : Huber loss
- dyn_wmse2 : Dynamic Weighted Mean Squared Error v2
- LPWAN : Low Power Wide Area Network
- PLR : Packet Loss Rate
- KMA : Korea Meteorological Administration
- ASOS : Automated Synoptic Observing System
- ISO : International Organization for Standardization
- EMS : Energy Management System

1. 서론

전력 수급균형을 위한 재생에너지의 시간 단위 정밀한 예측은 계획과 계통 운영의 불확실성을 완화하는 데 필수적인 요소로 인식되고 있다. 그 중 태양광 발전 및 기상 예측 분야에서 태양광 일사량은 발전량 산정, 전력망 안정화, 농업·건축 설계 등 다양한 의사결정의 핵심 입력 변수로 활용된다. 태양광 일사량은 태양 고도와 같은 천문학적 요인과 함께 구름의 형성·이동과 대기 중 수증기·에어로졸 농도 등 복합적인 대기 조건의 영향을 동시에 받기 때문에 본질적으로 높은 변동성을 갖는다.

선행 연구에 따르면 구름의 높이·분포와 이동 속도는 지표면에 도달하는 태양 복사량을 크게 변화시켜 일사량의 시

계열은 강한 비선형성과 무작위성을 띤다.^[1] 측정 단계에서도 비선형성과 무작위성에 대한 기술적 제약이 크게 작용하는데, 세계기상기구(WMO)는 일사량 측정 장비인 피라노미터를 2차 표준·1급·2급 등급(ISO 9060:2018의 Secondary Standard, First Class, Second Class)으로 분류하여^[2] 서로 다른 정확도를 가진 장비를 혼재하여 사용한다. 이러한 장비들은 주기적인 교정과 유지보수가 필요하고, 현장에서 관리가 쉽지 않아 데이터 품질이 일정하지 않을 뿐만 아니라, 태양복사 데이터에서도 통신 장애나 장비 고장으로 결측이 널리 나타난다.^[1]

일반적인 IoT 기반 기상·환경 센서는 주로 저전력·광역통신(LPWAN) 네트워크를 사용하는데, 이러한 네트워크는 전송률이 낮아 신호 손실이 흔하고 대규모 LoRaWAN 네트워크를 대상으로 한 연구에서는 대부분 장치의 패킷 손실률(PLR)이 0~10% 범위에 집중되어 있으며,^[3] 일부 장치만 매우 높은 PLR을 보였다. 또 다른 도시형 스마트 시티 사례에서는 20여 개의 장치가 3만 5천 개의 메시지를 송신한 결과 72.4%만이 데이터 서버에 성공적으로 수신되었고, 수신된 메시지의 99%는 10초 이내에 도달한 것으로 보고되었다.^[4] 이처럼 통신 오류로 인해 결측값이 무작위 혹은 주기적으로 발생하고, 한 번의 통신 장애가 연쇄적으로 데이터 손실을 확산시킨다. 환경 센서 결측을 사례별로 분석한 연구에서는 통신 오류가 원인인 결측을 무작위·주기적 두 종류로 구분하고, 센서 오류에 의한 결측은 측정값 급변 및 측정 범위 초과로 분류하였다.^[5]

또한, 센서 자체의 오류로 측정 범위를 벗어나는 급격한 값 변화나 전원 불안정, 기계적 마모, 외부 환경 요인(먼지·습기·물리적 충격 등)은 센서 정상 동작을 저해하여 결측이나 비정상 값을 유발한다.^[5] 야외에 설치된 기상 센서는 장기간 노출로 인해 성능 저하가 누적되며, 먼지 축적이나 습기 유입이 장비를 손상시켜 불안정한 데이터 수집을 초래한다. 기후 시계열을 대상으로 한 리뷰 연구에서는 강수량 자료의 결측률이 최대 41.8%에 달하는 사례도 보고되어, 실제 기상 데이터에서 결측 문제가 구조적으로 존재함을 시사한다.^[6]

마지막으로 단순 보간법으로 동적 특성을 가진 시계열의 패턴과 상관관계를 반영하지 못해 예측 성능이 저해된다.^[1] 결측치가 포함된 시계열 데이터에서 많이 쓰이는 방법은 선형

보간이나 전진 채우기(Last Observation Carried Forward, LOCF) 같은 단순 보간 기법을 적용하였으나 결측 구간이 짧고 값의 변화가 완만하다는 가정을 전제로 하며, 변동성이 큰 태양 복사량 시계열에는 적합하지 않다.

태양복사 자료를 대상으로 한 연구에서는 인접 관측소 간 거리 기준으로 태양 복사량을 상관관계로 보완하는 방법이 맑은 하늘이나 흐린 하늘에서는 비교적 잘 작동하지만, 부분적으로 흐린 하늘에서는 성능이 급격히 떨어지고 단일 값만 결측일 때에만 단순 보간이 효과적임을 보고하였다.^[1] 또한 여러 지점에서 다른 기상 변수를 활용해 복사를 예측하는 다중 선형 회귀나 차분거리 보간 등 전통적 기법은 결측률이 증가할수록 예측 정확도가 감소하는 한계를 보였다.^[1] 따라서 단순 보간만으로는 태양복사 시계열의 불규칙성과 대기 조건을 충분히 반영할 수 없으며, 장기간의 결측에는 더 정교한 기법이 요구된다. 이러한 한계를 보완하기 위해, 초기 연구에서는 순환 신경망이나 GAN을 활용하여 시계열 결측치를 모델 내부에서 직접 복원하는 접근이 제안되었다.^[7~9] 이러한 방법들은 결측 구간의 연속성을 일정 부분 회복하는 데 효과적이었으나, 결측을 복원한 이후 예측을 수행하는 구조를 따른다는 공통적인 특징을 가진다. 최근에는 순환 신경망, attention, diffusion 모델을 활용하여 시계열 결측치를 직접 복원하거나 생성하는 연구들이 활발히 제안되고 있다.^[10~14] 이러한 접근은 결측 복원 정확도를 향상시키는 데 효과적이거나, 모델 구조 복잡도와 계산 비용이 증가하며, 실시간 예측 시스템에 적용하기에는 제약이 존재한다. 고신뢰 태양광 예측 모델을 구축하기 위해서는 연속적이고 신뢰할 수 있는 시계열 데이터가 필요하지만, 통신·센서 오류, 장비 노후 등의 요인으로 장기간 결측치가 발생하는 경우가 빈번하다. 해당 문제로 인하여 태양복사 데이터는 기상·대기 조건에 따른 본질적 불확실성과 측정 장비의 물리적·운영적 한계를 동시에 내포한다.

이러한 한계를 극복하기 위해 최근 연구들은 적응적 학습(adaptive learning) 모델을 통해 결측치를 직접 학습에 포함하거나 결측 패턴을 고려한 모델링을 수행한다. 예를 들어 통신 오류·센서 오류 등 결측 발생 유형을 분류해 각각에 적합한 보간·모델을 적용하는 접근도 제안되었다.^[5] 본 연구는 이러한 배경을 바탕으로 태양광 일사량 예측에서 결측의 발생 빈도와 연속 길이를 파악하고, 결측 구간

의 불확실성을 학습 과정에 반영하는 동적 가중 손실 함수 DYN_wMSE2를 제안한다. 이를 통해 결측이 존재하는 상태에서도 안정적인 예측 성능을 확보할 수 있음을 실험적으로 검증한다.

2. 본론

2.1 데이터 출처 및 데이터 분석

본 연구에서는 대한민국 기상청이 운영하는 기상자료 개방포털에서 제공하는 종관 기상관측(Synoptic Meteorological Observation) 자료를 활용하였다. 종관 기상관측 자료는 전국에 분포된 관측소에서 일정한 시간 간격으로 기온, 강수량, 풍속, 습도, 기압 등 주요 기상 요소를 표준화된 방식으로 관측·기록한 자료로, 장기간 축적된 시계열 특성과 공간적 비교 가능성을 동시에 갖는다는 점에서 기초 분석 자료로서의 활용도가 높다.

본 절에서는 사용된 기상자료의 출처와 구성 요소를 명확히 기술하고, 이후 분석의 기반이 되는 데이터 특성을 정리한다. 특히 본 연구의 목적은 단순한 기상 현상 설명이 아니라, 예측 모델 및 분석 기법 적용을 위한 데이터 품질과 구조적 특성을 검토하는 데 있으므로, 관측 자료의 신뢰성과 한계를 함께 고려한다.

실제 기상 관측 환경에서는 관측 장비의 오작동, 유지보수 과정에서의 측정 공백, 기상 조건에 따른 센서 감도 저하, 통신 장애 등으로 인해 결측치(missing values)가 불가피하게 발생한다. 이러한 결측치는 시계열 연속성을 저해하고, 통계적 분석이나 학습 기반 모델의 성능에 직접적인 영향을 미칠 수 있다. 따라서 결측치의 존재 자체를 단순한 오류로 간주하기보다는, 발생 원인과 형태를 데이터 특성의 일부로 해석할 필요가 있다.

본 연구에서는 종관 기상관측 자료 중 전주와 고창 지역 관측소 데이터 사례로, 결측치의 발생 빈도, 지속 시간, 시간적 분포 특성을 분석한다. 이를 통해 결측치가 무작위적으로 발생하는 경우와 특정 구간에 집중되는 경우를 구분하고, 결측치 발생 패턴에 따라 데이터 특성을 분류하기 위한 가설을 제시한다. 이러한 분석 결과는 이후 결측치 처리 전략 수립과 예측 모델 설계 단계에서 기초 가정으로 활용된다.

2.2 전주·고창 관측소 결측치 패턴 분석

본 절에서는 전주와 고창 지역 중관 기상관측소의 자료를 대상으로 각각 최근 3년간의 시계열 데이터를 분석하여, 결측치의 시간적 발생 양상을 비교·분석한다. 두 관측소 모두 동일한 관측 체계와 자료 형식을 따르지만, 지역적 환경과 관측 조건의 차이로 인해 결측치의 분포 양상에는 차이가 존재한다.

Fig. 1과 같이 결측치 분포를 구간 평균 기반 히트맵으로 시각화한 결과, 전주와 고창 데이터 모두에서 결측이 단발적으로 발생하는 구간과, 연속된 시간 구간에 걸쳐 지속적으로 나타나는 구간이 명확히 구분됨을 확인할 수 있었다.

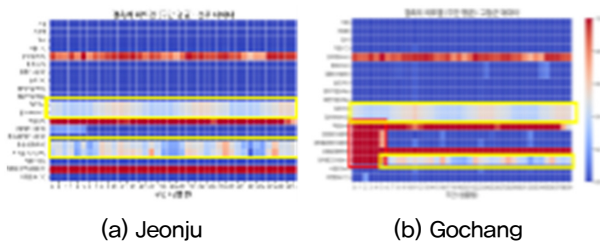


Fig. 1. Heatmap of missing values (Interval average)

전자의 경우 특정 시점에서 일시적으로 발생하는 누락으로, 통신 지연이나 일시적인 센서 이상과 같은 요인에 의해 발생했을 가능성이 높다. 반면 후자는 일정 기간 동안 동일 기상 요소에서 연속적 및 광범위하게 결측이 나타나는 형태로, 센서 고장이나 장비 교체·점검과 같은 구조적 요인에 기인했을 가능성을 시사한다.

이러한 분석 결과는 결측치를 하나의 균질한 오류로 취급하는 접근이 데이터의 실제 생성 과정을 충분히 반영하지 못함을 보여준다. 즉, 전주와 장수 관측소의 3년 치 자료 분석을 통해 결측치를 일시적인 누락과 장기적인 누락으로 구분할 수 있음을 확인하였으며, 이는 결측치 처리 방법을 설계할 때 시간적 지속성과 발생 패턴을 함께 고려해야 함을 의미한다.

2.3 전처리 전략에 대한 논의 및 손실 함수 설계로의 전환

앞선 결측치 패턴 분석 결과는, 결측치가 단순히 제거되거나 보정되어야 할 오류가 아니라 시계열 데이터 생성 과

정의 불필요한 특성임을 보여준다. 특히 일시적 누락과 장기적 누락이 혼재된 상황에서는, 결측 구간에 대해 동일한 전처리 규칙을 적용하면 결측의 시간적 구조와 정보 신뢰도의 차이가 소거될 수 있다.

이에 본 연구에서는 결측치 보정을 위한 복잡한 전처리 과정을 최소화하고, 관측 데이터의 원형 특성과 결측 발생 양상을 가능한 한 보존하는 방향을 채택하였다. 기존 시계열 결측 처리 연구들은 주로 결측값을 복원하거나 조건부 생성하여 입력 데이터를 보완하는 방식에 초점을 두고 있다.^[15] 반면 본 연구는 결측값을 복원 대상으로 취급하기보다, 결측 발생에 따른 정보 신뢰도의 변화를 학습 과정에서 직접 반영하도록 손실 함수 차원에서 접근한다. 본 연구는 결측을 복원 대상이 아닌 정보 신뢰도의 변화로 해석하고, 이를 손실 함수 설계에 직접 통합함으로써 결측 환경 변화에 따른 예측 성능 저하를 구조적으로 완화하고자 한다. 이는 전처리의 중요성을 부정하기 위함이 아니라, 전처리 단계에서 해결하기 어려운 결측의 시간적·구조적 특성을 학습 과정에서 직접 반영하기 위한 설계적 선택이다.

시계열 기상 데이터에서는 입력 변수들의 예측 기여도가 시간, 계절, 기상 상황에 따라 지속적으로 변화하며, 결측이 장기적으로 지속되는 구간에서는 일부 변수의 정보 신뢰도가 현저히 저하된다. 그러나 평균제곱오차(MSE)나 가중 평균제곱오차(Weighted Mean Squared Error, WMSE)와 같이 고정된 가중치를 사용하는 정적 손실 함수는 이러한 변화하는 특성과 결측 환경을 충분히 반영하는 데 한계를 가진다.

따라서 본 연구에서는 결측 발생 여부와 입력 특성의 중요도 변화가 학습 과정에 직접 반영될 수 있도록, 정적 손실 함수가 아닌 시계열 환경과 결측 특성에 따라 동적으로 작동하는 손실 함수 설계로 분석의 초점을 전환한다. 이러한 접근은 전처리 의존도를 낮추는 동시에, 실제 관측 환경에서 발생하는 비정상성과 결측 문제에 보다 강건한 예측 모델을 구축하기 위한 기반을 제공한다.

2.4 동적 손실 함수를 포함한 예측 시스템 구조

Fig. 2는 본 연구에서 제안하는 예측 시스템 전체 구조를 나타낸다. 시스템은 최소화된 전처리 단계, 시계열 예측 모델, 그리고 결측 발생 특성과 입력 변수의 중요도 변화를

반영하여 동적으로 작동하는 손실 함수로 구성된다. 이러한 구조는 결측치를 사전에 완전히 제거하거나 보정하는 대신, 학습 과정 전반에 걸쳐 결측 정보와 신뢰도를 함께 고려하도록 설계하였다.

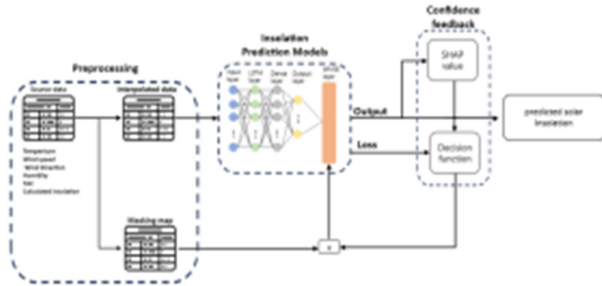


Fig. 2. Deep learning-based solar insolation prediction with SHAP-driven reliability assessment

전처리 단계에서는 원시 관측 데이터를 기반으로 필수적인 정합 처리만을 수행하며, 결측 발생 여부는 마스킹 맵(masking map)의 형태로 별도로 유지된다. 이는 결측 구간에 인위적인 연속성을 부여하지 않으면서도, 결측 정보가 이후 학습 단계에서 명시적으로 활용될 수 있도록 하기 위함이다.

예측 모델은 시계열 특성을 학습할 수 있는 순환 신경망 기반 구조를 중심으로 구성되며, 입력층과 은닉층을 거쳐 출력된 예측값은 손실 함수 계층으로 전달된다. 이때 손실 계산 과정에서는 단순한 예측 오차뿐만 아니라, 마스킹 정보를 통해 결측 구간과 비결측 구간을 구분하고, 각 구간의 정보 신뢰도를 고려한 가중 오차가 산출된다.

또한 본 시스템에서는 학습 과정에서 도출되는 입력 변수의 기여도를 정량화하기 위해 SHAP 값을 활용한다. SHAP 기반 중요도 정보는 예측 결과에 대한 신뢰도 피드백으로 해석되며, 이는 손실 함수 내 의사결정 함수(decision function)에 반영되어 가중치 조정에 사용된다. 이를 통해 손실 함수는 학습이 진행됨에 따라 입력 특성의 상대적 중요도 변화와 결측 환경에 적응적으로 반응하게 된다.

이와 같은 동적 손실 함수를 포함한 예측 시스템 구조는, 시계열 데이터의 비정상성과 결측 문제를 전처리 단계에서 단순화하기보다 학습 과정 내부로 통합하여 다루는 접근이다. 이러한 설계는 이후 장에서 제시하는 손실 함수의 수식

적 정의와 학습 알고리즘 설명을 위한 개념적 기반으로 활용된다. 한편 MICE와 같은 다중 대체 기반 결측치 보간 기법은 정적 데이터 환경을 전제로 설계되어, 결측값을 사전 보정한 후 예측을 수행하는 방식이 일반적이다. 이러한 접근은 결측 복원 정확도 측면에서는 효과적일 수 있으나, 결측 발생 자체를 학습 신호로 활용하거나 예측 과정에서의 신뢰도 변화를 반영하기는 어렵다. 반면 본 연구의 방법은 결측 정보를 손실 함수 수준에서 직접 반영함으로써, 결측 환경 변화에 따른 예측 성능 저하를 완화하는 데 초점을 둔다.

2.4.1 문제 설정 및 동적 손실 함수의 수식적 정의

본 연구에서 다루는 문제는 다변량 기상 시계열 데이터를 이용하여 단일 시점의 태양 일사량을 예측하는 회귀 문제로 정의된다. 각 시점 i 에 대해 입력 데이터는 길이 L 의 시간 윈도우와 F 개의 기상 변수로 구성된 행렬로 표현된다.

$$X^{(i)} \in \mathbb{R}^{L \times F} \quad (1)$$

연구에서는 1시간 단위 기상 관측 데이터를 사용하므로, $L=24$ 로 설정하면 과거 24시간의 일주기(diurnal cycle) 패턴을 입력 컨텍스트로 완전히 포함할 수 있다. 태양 일사량은 일출·일몰을 중심으로 24시간 주기의 구조적 변동성을 가지므로, 하나의 완전한 일주기를 lookback 길이로 설정하는 것은 일사량 예측 선행 연구에서 공통적으로 사용되는 기준이다.^[1,6] L 을 24보다 크게 설정하면 복수의 일주기가 포함되어 모델 입력이 과거 패턴에 희석될 수 있으며, 연산 비용과 배치 처리 효율 측면에서도 불리하다. 반대로 L 이 지나치게 작으면 일사량의 단기 변동성과 결측 패턴이 충분히 반영되지 않는다. 이러한 근거를 바탕으로 $L=24$ 를 고정값으로 적용하며, 이 설정은 3.3절 Table 1에도 명시하였다.

입력 데이터에서 결측 여부를 명시적으로 반영하기 위해, 시간 인덱스 t 와 feature 인덱스 f 에 대해 결측 마스크를 다음 식 (2)와 같이 정의한다.

$$M_{t,f} = \begin{cases} 1, & \text{if } X_{t,f} \text{ is missing} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

식 (2)는 입력 행렬 내 각 원소의 결측 여부를 이진 변수로 표현한 것으로, 이후 결측 기반 신뢰도 계산의 기초로 사용된다. 결측이 존재하는 위치에서는 더 높은 불확실성이 존재한다고 가정하여, 결측 마스크를 기반으로 한 가중 맵을 다음 식 (3)과 같이 정의한다.

$$W_{t,f} = 1 + \alpha \cdot M_{t,f}, \alpha > 0 \quad (3)$$

식 (4)에서 α 는 결측에 대한 페널티 계수로, 결측이 발생한 위치의 가중치를 증가시키는 역할을 한다. 이후 전체 데이터에 대해 평균 가중치가 1이 되도록 정규화를 수행한다. 각 입력 윈도우 i 에 대해 결측 분포를 하나의 스칼라 값으로 요약하기 위해, 모든 시간-feature 축에 대해 평균을 취한 윈도우 단위 신뢰도 계수를 정의한다.

$$r^{(i)} = \frac{1}{LF} \sum_{t=1}^L \sum_{f=1}^F W_{t,f}^{(i)} \quad (4)$$

식 (4)는 해당 샘플이 포함하는 결측의 상대적 정도를 나타내며, 값이 클수록 결측이 많이 포함된 입력임을 의미한다. 이 값 역시 데이터 전반의 평균으로 정규화되어 상대적인 신뢰도 지표로 사용된다.

$$\Phi_f = \frac{1}{L} \sum_{t=1}^L |\phi_{t,f}| \quad (5)$$

식 (5)는 각 feature가 전체 예측 과정에 기여한 평균적인 중요도를 나타낸다. 이후 모든 feature에 대해 평균이 1이 되도록 정규화를 수행한다. 이후 SHAP 값의 과도한 집중을 방지하기 위해 다음과 같은 비선형 압축 함수를 적용한다.

$$\Phi_f^{reg} = 1 + \frac{\tilde{\Phi}_f - 1}{1 + \lambda|\tilde{\Phi}_f - 1|} \quad (6)$$

SHAP 값은 특정 feature에 중요도가 과도하게 집중되는 경향이 있으며, 이를 그대로 손실 가중치로 사용하면 학습 초반에 가중치 분산이 커져 학습 불안정이 발생할 수 있

다. 식 (6)의 비선형 압축 함수는 중요도가 큰 feature의 가중치를 완화하여 분포를 평탄화하는 역할을 한다. λ 가 클수록 중요도 분포가 균일화되고, $\lambda=0$ 이면 원래 SHAP 분포를 그대로 사용한다. 본 연구에서는 $\lambda=0.5$ 를 적용하였으며, 이는 3.4절 Table 3의 하이퍼파라미터 그리드 탐색에서 검증 손실 기준으로 선정된 값이다. $\lambda=0.5$ 는 특정 feature의 중요도 독점을 방지하면서도 상대적인 중요도 차이를 손실 함수에 의미 있게 반영할 수 있는 균형점으로 해석된다.

결측 기반 가중 맵과 SHAP 기반 중요도를 결합하여, feature 인지적 마스크 가중치를 다음과 같이 정의한다. 이렇게 결측이 많고 동시에 예측에 중요한 feature가 손실될 때 더 큰 페널티가 부여되는 결합 가중치를 만들어낸다. 이후 전체 평균으로 정규화를 수행하며, 이를 다시 윈도우 단위로 평균하여 샘플별 최종 가중치를 정의한다.

$$w^{(i)} = \frac{1}{LF} \sum_{t=1}^L \sum_{f=1}^F \hat{W}_{t,f}^{(i)} \quad (7)$$

식 (7)은 각 학습 샘플의 전반적인 신뢰도 보정 계수를 나타낸다. 여기에 조절 계수 β 를 적용하여 동적 가중 강도를 조정한다. 이를 샘플별 동적 가중치로 확장한 동적 가중 평균제곱오차 손실 함수(dyn_wmse2)는 다음과 같이 정의된다.

$$\mathcal{L}_{dyn-wmse2}(\theta) = \frac{1}{N} \sum_{i=1}^N \tilde{w}^{(i)} \ell^{(i)} \quad (8)$$

식 (8)은 결측 신뢰도, SHAP 기반 특성 중요도, 그리고 결정 계수를 결합한 동적 가중 손실 함수로, 단순한 오차 최소화를 넘어 결측 환경과 특성 중요도를 학습 과정에 직접 반영하는 결정 함수로 해석될 수 있다.

2.4.2 비교 손실: grad×input 기반 중요도

본 절에서는 제안한 SHAP 기반 동적 손실 함수(dyn_wmse2)의 필요성을 검증하기 위한 비교 기준으로서, 기울기 (gradient)에 기반한 단순 attribution 지표인 grad×input 을 이용한 손실 함수 변형을 정의한다. 본 비교 손실은 단순히 동적 가중치의 존재 여부를 확인하기 위한 것이 아니

라, 동적 손실 설계에서 어떤 형태의 중요도 정의가 효과적 인가를 분석하기 위한 대조군으로 설정된다.

시계열 입력 윈도우 하나는 길이 L 과 전체 입력 채널 수 F_t 를 갖는 행렬 $X^{(i)} \in \mathbb{R}^{L \times F_t}$ 로 정의한다. 본 연구에서는 수치 기상 변수 F 에 더해 동일 차원의 결측 마스크 채널을 결합하므로 $F_t = 2F$ 가 된다. 예측 모델은 ANN 또는 LSTM 기반의 회귀 함수 $f_\theta(\cdot)$ 로 표현되며, 출력은 스칼라 예측 값 $\hat{y}^{(i)} = f_\theta(X^{(i)})$ 이다.

입력 성분 $X_{t,f}^{(i)}$ 에 대한 출력의 변화율 $\partial \hat{y}^{(i)} / \partial X_{t,f}^{(i)}$ 는 해당 성분의 국소적 민감도를 나타낸다. 본 비교 손실에서는 가장 단순한 1차 근사 기반 attribution으로 다음과 같은 식 (12)로 중요도를 정의한다.

$$A_{t,f}^{(i)} = \left| \frac{\partial \hat{y}^{(i)}}{\partial X_{t,f}^{(i)}} \right| \cdot X_{t,f}^{(i)} \quad (9)$$

위 정의는 1차 테일러 근사인 다음 식의 $f_\theta(X + \Delta X) \approx f_\theta(X) + \sum_{t,f} (\partial f_\theta / \partial X_{t,f}) \Delta X_{t,f}$ 에서 $\Delta X_{t,f} = X_{t,f}$ 로 두는 단순화에 기반하며, grad×input을 각 입력 성분의 국소 선형화된 기여도에 대한 휴리스틱 지표로 해석할 수 있게 한다.

손실 가중치로 활용하기 위해 $A_{t,f}^{(i)}$ 는 시간-채널 축에 대해 평균되어 샘플 단위 스칼라 중요도 $I^{(i)}$ 로 축약된다.

$$I^{(i)} = \frac{1}{LF} \sum_{t=1}^L \sum_{f=1}^{F_t} A_{t,f}^{(i)} \quad (10)$$

식 (10)은 학습 집합 전체에서 평균이 1이 되도록 정규화 되어 상대적 중요도 지표 $\tilde{I}^{(i)}$ 로 사용된다. 이후 2.4.1절에서 정의한 결측 기반 샘플 가중치 $w_{mask}^{(i)}$ 와 결합하여, 다음과 같은 비교용 동적 가중치를 구성한다.

$$\tilde{w}_{grad}^{(i)} = 1 + \beta (w_{mask}^{(i)} \cdot \tilde{I}^{(i)} - 1) \quad (11)$$

식 (11)을 이용한 비교 손실 함수는 다음과 같이 정의된다.

$$\mathcal{L}_{dym-wmse(mask+input \times grad)}(\theta) = \frac{1}{N} \sum_{i=1}^N \tilde{w}_{grad}^{(i)} (\hat{y}^{(i)} - y^{(i)})^2 \quad (12)$$

본 비교 손실은 결측량과 중요도를 결합한 동적 가중 손실이라는 점에서는 dyn_wmse2와 동일한 구조를 갖지만, 중요도 정의가 SHAP이 아닌 국소 기울기 기반 휴리스틱에 의존한다는 점에서 본질적인 차이를 가진다. 이에 따라 본 연구에서는 다음의 ‘가설’을 설정한다.

“동적 가중 손실의 성능 향상은 단순히 가중치가 학습 과정에서 변화하기 때문이 아니라, 예측 기여도를 안정적이고 전역적으로 반영할 수 있는 중요도 정의(SHAP 등)의 사용 여부에 의해 좌우된다.”

가설은 이후 실험에서 SHAP 기반 dyn_wmse2와 grad×input 기반 비교 손실의 성능, 안정성, 결측 비율 증가에 따른 거동을 비교·분석함으로써 검증되고, 이를 통해 단순한 동적 손실 설계를 넘어, 시계열-결측 환경에서 요구되는 attribution의 성질을 실증적으로 분석한다.

2.4.3 SHAP 계산 오버헤드 및 실시간 EMS 적용 가능성

실시간 EMS 환경에서의 적용 가능성과 관련하여, SHAP 기반 가중치 갱신은 학습 단계에서만 수행된다. 추론 단계에서는 학습 완료 시점의 SHAP 기반 중요도를 고정값으로 사용하므로, 실시간 추론에서 추가적인 SHAP 계산 비용이 발생하지 않는다.

데이터 분포 변화에 대한 적응 전략으로는, 롤링 윈도우 방식의 재학습 후 모델 교체를 적용할 수 있다. 이 방식은 최신 N 일 관측 데이터로 새 모델을 오프라인 학습하고, 검증 손실 기준으로 현재 배포 모델보다 우수할 경우 교체하는 구조이다. 이때 중요한 점은, SHAP 기반 feature 중요도는 단기 관측 노이즈가 아닌 기후적 구조에 의존하므로 매우 천천히 변화하는 특성을 가진다. 본 실험에서도 학습 후반부 epoch 간 SHAP 분포 변화폭은 수렴 이후 안정적으로 유지되었다. 따라서 롤링 재학습으로 교체된 모델의 SHAP 분포는 이전 모델과 크게 다르지 않으며, 모델 교체 시 예측 거동의 급격한 불연속성 없이 EMS 운영 안정성을 확보할 수 있다.

3. 실험

3.1 실험 데이터 세트

본 장에서는 2장에서 정의한 동적 손실 함수 및 비교 손

실의 효과를 검증하기 위해, 전주(ASOS) 종관관측 자료를 기반으로 실험을 수행한다. 실험 데이터는 전주 관측소의 2023년 및 2024년 시계열 자료를 사용하며, 기상청(KMA) 기상자료개방포털에서 제공되는 동일 형식의 관측 데이터를 연도별로 수집한 뒤 통합하여 학습·검증·평가에 활용한다.

입력 변수는 기온, 강수량, 풍속, 풍향, 습도, 기압(현지/해면), 일조, 적설, 운량(전운량/중하층운량), 시정 등 표준 관측 항목으로 구성되며, 목표 변수(target)는 태양 일사량(Solar Radiation)이다. 전처리 과정에서 일부 범주형, 코드성 변수는 예측 입력에서 제외하고, 연속 수치형 기상 변수 중심으로 모델을 구성하였다.

본 연구의 핵심은 결측치를 전처리 단계에서 과도하게 보정하기보다, 결측 발생 정보(마스크)를 학습 과정에 포함하여 손실 함수가 이를 반영하도록 설계하는 데 있다. 따라서 결측이 포함된 원본 관측값은 결측 여부를 유지한 채로 마스크를 구축하고, 모델 입력에는 결측 위치를 0으로 채운 수치 행렬과 결측 마스크 채널을 결합한 입력을 사용한다(입력 채널 수는 수치 변수 F 와 마스크채널 F 를 결합하여 $F_i = 2F$). 또한 입력 수치 행렬은 StandardScaler로 정규화하고, 타겟 변수(일사량) 역시 StandardScaler로 정규화하여 학습 안정성을 확보하였다.

3.2 결측 주입 시나리오 적용 및 실험 설계

2.2절에서 확인한 바와 같이 결측은 단발성 누락과 구간형 누락이 혼재되어 나타나므로, 본 실험에서는 결측 생성 방식을 다음의 두 가지 시나리오로 구분하여 실험을 수행한다.

첫 번째로 랜덤 결측(Random Missing) 시나리오로 전체 시계열 전 구간에 걸쳐 결측이 무작위로 산재하도록 결측을 주입한다.

다음 두 번째로 블록 결측(Block/Continuous Missing) 시나리오로 특정 시간 구간에 결측이 연속적으로 발생하도록(센서 다운타임/통신 장애와 유사) 결측을 주입한다.

두 시나리오 모두 결측 주입 비율은 0~40%까지로 설정하며, 각 실험에서 목표 결측 비율까지 랜덤하게 결측을 생성한다.

블록 결측은 연속 구간 길이 최소 3에서 최대 12까지의 블록 랜덤 노이즈를 반복적으로 생성하여 결측을 구성하

며, 본 구현에서는 동일한 결측 비율에서도 연속 결측의 영향을 반영하기 위해 블록 결측 비중을 설정한다. 이후 남은 결측 수는 랜덤 위치에 추가 주입하여 최종적으로 목표 결측 비율을 정확히 맞춘다.

3.3 모델 및 학습 설정

본 절에서는 실험의 재현 가능성을 위해 모델 학습 설정과 손실 함수 탐색 그리드(grid) 구성을 명시한다. 공통 설정은 Table 1과 같다.

비교 모델은 Table 2와 같이 우선 ANN과 LSTM 두 계열로 구성하며, 동일한 입력(수치+마스크 결합)과 동일한 시간 분할 규칙을 적용하여 손실 함수 차이에 따른 성능 변화를 비교한다.

본 연구는 연속 값을 예측하는 회귀 문제로 정의되며, 회귀 문제에서 가장 표준적으로 사용되는 평가 지표인 평균 제곱오차(MSE)를 주요 성능 지표로 채택한다. MSE는 예측 오차를 제공하여 계산하므로 큰 예측 오차에 대해 민감하게 반응하는 특성을 가지며, 이는 예측 결과의 안정성과 이상 오차 발생 여부를 평가하는 데 적합하다. 특히 기상 및 에너지 예측과 같이 단일 시점에서의 큰 오차가 시스템

Table 1. Hyperparameter settings for model training and data splitting

Parameter	Symbol/parameter	Data
Input Window Length	L	24
Prediction Horizon	H	1
Data (test)	TEST_SIZE	0.15
Data (Validation)	VAL_SIZE	0.15
Batch size	BATCH	128
Maximum epochs	EPOCHS	40
Early stopping	PATIENCE	6
Random seed	SEED	42

Table 2. Model-specific hyperparameters for ANN and LSTM

Model	Main hyperparameters
ANN	hidden=(512, 256, 128, 64, 32), dropout=0.05, BatchNorm
LSTM	units=96, bidirectional=True, dropout=0.1
Equally	adapted temporal smoothing($\alpha = 0.7$) Mask channel combination

운영 및 의사결정에 직접 영향을 미칠 수 있는 문제 설정에서는, 평균적인 오차 크기뿐만 아니라 극단적인 오차의 발생 여부를 함께 고려하는 것이 중요하다.

이러한 특성으로 인해 MSE는 모델의 실질적인 신뢰도와 위험도를 동시에 반영할 수 있는 평가 지표로 활용된다. 이에 따라 이후의 모든 성능 비교 및 분석은 MSE를 기준으로 수행한다.

3.4 dyn_wmse2 하이퍼파라미터 그리드 및 활성화 조건

본 연구에서 제안하는 dyn_wmse2(mask+SHAP) 손실 함수는 학습 전 구간에서 항상 동일하게 적용되는 정적 손실이 아니라, 학습 진행 상태에 따라 SHAP 기반 가중치 선택적으로 활성화되는 구조를 갖는다. 따라서 손실 함수의 동작 특성과 민감도를 체계적으로 분석하기 위해, 표3과 같은 하이퍼파라미터 그리드를 설정하여 실험한다. 동적 손실의 활성화 조건은 사전에 설정한 임계값 τ 이하로 검증 손실(val_loss)이 감소하는 시점으로 정의한다. 이는 모델이 기본적인 시계열 패턴을 충분히 학습하지 못한 초기 단계에서, 불안정한 중요도 추정이 손실 계산에 과도하게 개입하는 것을 방지하기 위한 설계이다. 즉, 학습 초반은 기본 MSE 기반 학습을 유지하고, 일정 수준 이상의 예측 안정성이 확보된 이후에만 SHAP 기반 가중 업데이트를 허용한다.

활성화 이후에는 매 epoch마다 SHAP 기반 feature 중요도를 재추정하고, 이를 결측 신뢰도 정보와 결합하여 샘플

가중치를 갱신한다. 이 과정에서 손실 함수는 단순한 오차 누적 함수가 아니라, 학습 상태와 입력 정보의 신뢰도에 따라 조정되는 결정 함수로 동작하게 된다.

실험 과정에서는 이러한 손실 함수의 동작 상태를 추적하기 위해 dyn_mode 변수를 함께 기록한다. dyn_mode는 손실 함수의 상태를 비활성(Mode 0), 초기 활성화(Mode 1), 지속 갱신(Mode 2)으로 구분하여 나타내며, 이를 통해 동적 손실 함수가 언제 개입되었고, 이후 학습 과정에서 어떤 방식으로 유지·조정되었는지를 분석할 수 있도록 한다.

3.5 성능 비교 결과

본 절에서는 제안한 dyn_wmse2(mask+SHAP)의 동적 손실 함수 성능을 기존 손실 함수들과 비교·분석한다.

비교는 결측 주입 비율에 따라 수행되며, 먼저 랜덤 결측 시나리오에 대한 결과를 제시한 후, 다음 절에서 블록 결측 시나리오로 확장한다. 모든 결과는 테스트 셋 기준이며, 성능 평가는 앞서 정의한 바와 같이 MSE를 기준 지표로 사용하여 Table 4와 같이 순위를 산정하였다.

3.5.1 랜덤 결측 시나리오 실험 결과

랜덤 결측 시나리오는 전체 시계열 구간에 걸쳐 결측이 무작위로 산재하는 상황을 가정한다.

이는 통신 지연이나 일시적인 센서 오류가 불규칙적으로 발생하는 환경을 모사한다. 본 실험에서는 결측 비율을 0.0부터 0.4까지 단계적으로 증가시키며, 각 결측 비율에 대해 손실 함수별 성능을 Table 5와 같이 비교하였다.

Table 3. Parameter settings of the proposed dynwmse2 loss function

Parameter	Symbol	Value	Remarks
Dynamic weighting strength coefficient	β	{0.0, 0.1, ..., 1.0}	Strength of influence of SHAP missing-data-based sample weights
Activation threshold	τ	{0.08, 0.10}	Dynamic weighting update start criteria
Importance stabilization coefficient	λ_{reg}	0.5	Suppression of excessive concentration in SHAP importance distribution

Table 4. Random Missing Performance ANN

MSE loss	Ratio				
	0.0	0.1	0.2	0.3	0.4
dyn_wmse2 (mask+SHAP)	0.0590	0.0779	0.1001	0.1158	0.1276
mae	0.0624	0.098	0.1127	0.1418	0.1625
dyn_wmse (mask+grdxinput)	0.0885	0.1267	0.1541	0.1645	0.2348
mse	0.0919	0.1298	0.1389	0.2088	0.199
huber	0.0919	0.1026	0.1123	0.1517	0.1797
wmse(mask)	0.2831	0.2907	0.2453	0.2788	0.2988
wmse(mask*shap)	0.3275	0.2554	0.3309	0.3524	0.3409

Table 5. Random missing performance LSTM

MSE	Ratio				
	0.0	0.1	0.2	0.3	0.4
loss	0.0	0.1	0.2	0.3	0.4
dyn_wmse2 (mask+SHAP)	0.0420	0.0642	0.0825	0.1047	0.1023
mae	0.0461	0.0840	0.0861	0.1082	0.1194
dyn_wmse (mask+grdxinput)	0.0501	0.0743	0.1062	0.1154	0.1293
mse	0.0540	0.0872	0.1041	0.1188	0.1386
huber	0.0491	0.0841	0.1070	0.1078	0.1450
wmse(mask)	0.3519	0.3547	0.2768	0.3109	0.2876
wmse(mask*shap)	0.3333	0.2679	0.3162	0.3768	0.3137

3.5.2 성능 비교 결과에 대한 종합 분석

랜덤 결측 시나리오에 대한 ANN 및 LSTM 실험 결과를 종합하면, 모든 결측 비율(ratio = 0.0~0.4)과 모든 모델 구조에서 dyn_wmse2(mask+SHAP)이 일관되게 MSE 기준 1위를 차지함을 확인할 수 있다. 이는 단일 특정 조건에서의 우수성이 아니라, 결측 환경 전반에 걸친 구조적 우위를 의미한다.

특히 본 연구에서는 단순히 “가장 낮은 MSE를 기록했다”라는 결과 자체보다, 결측 비율이 증가함에 따라 성능이 얼마나 완만하게 저하되는지, 즉 성능 방어 능력(robustness)에 주목한다. 이를 위해 결측 비율 증가에 따른 절대 MSE 변화 양상을 시각화하여 Fig. 3처럼 분석하였다.

ANN 기반 실험 결과를 결측 비율 대비 절대 MSE 곡선으로 비교하면, dyn_wmse2(mask+SHAP)는 가장 완만한 기울기를 유지하며 결측 비율이 증가함에 따라 성능이 점진적으로 저하되는 양상을 보인다. 반면, MSE, MAE, Huber와 같은 정적 손실 함수들은 결측 비율이 0.2를 초과하는 구간부터 오차 증가 폭이 급격히 확대된다.

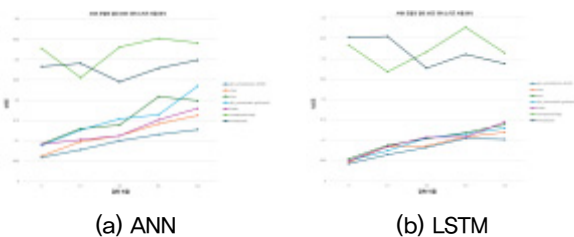


Fig. 3. Performance degradation comparison under increasing missing-data ratios

정적 WMSE 계열(wmse(mask), wmse(mask × shap))은 결측 비율 증가에 따라 비선형적으로 MSE가 급등하는 현상을 보이며, 이는 결측 정보를 정적으로 기중하는 방식이 오히려 학습 안정성을 저해할 수 있음을 시사한다. dyn_wmse(mask+grad × input)는 기본 MSE 대비 일정 수준의 방어 효과를 보이지만, 결측 비율이 높아질수록 dyn_wmse2와의 격차가 점차 확대된다.

이러한 결과는 ANN 구조와 같이 시간적 의존성이 제한적인 모델에서조차, SHAP 기반 특성 중요도를 결합한 동적 손실 함수가 결측 환경 변화에 가장 안정적으로 대응함을 의미한다. 또한 LSTM 모델에서도 동일한 경향이 관찰된다. dyn_wmse2(mask+SHAP)는 모든 결측 비율 구간에서 최저 MSE를 유지하며, 특히 결측 비율이 0.3~0.4로 증가하는 고결측 환경에서도 성능 저하 폭이 가장 작다. 이는 LSTM의 시계열 표현력이 결측 환경 변화에 취약해지는 상황에서도, 손실 함수 차원에서의 동적 보정이 효과적으로 작동했음을 의미한다.

주목할 점은, LSTM에서는 dyn_wmse(mask+ grad × input)가 ANN 대비 상대적으로 더 나은 성능을 보이는 구간도 존재하지만, 모든 결측 비율에서 dyn_wmse2를 추월하지는 못한다는 점이다. 이는 국소 기울기 기반 attribution이 시계열 비선형성과 결측 패턴 변화에 대해 제한적인 정보를 제공하는 반면, SHAP 기반 중요도는 입력 변수 간 상호작용과 전역적 기여도를 보다 안정적으로 반영하기 때문으로 해석할 수 있다.

Table 6과 같이 ANN과 LSTM 결과를 종합하면, dyn_wmse2(mask+SHAP)는 결측 비율이 증가함에도 불구하고

Table 6. Block missing performance ANN

MSE	Ratio				
	0.0	0.1	0.2	0.3	0.4
loss	0.0	0.1	0.2	0.3	0.4
dyn_wmse2 (mask+SHAP)	0.0448	0.0594	0.0659	0.0751	0.0852
mae	0.0640	0.0750	0.1120	0.1180	0.1410
dyn_wmse (mask+grdxinput)	0.0830	0.0850	0.1320	0.1310	0.1530
mse	0.0820	0.0880	0.1250	0.1230	0.1730
huber	0.0840	0.0860	0.1450	0.1470	0.1790
wmse(mask)	0.2170	0.2270	0.2490	0.3910	0.3260
wmse(mask*shap)	0.1790	0.2930	0.6000	0.3580	0.4100

고 성능 저하율($\Delta MSE / \Delta ratio$)이 가장 낮은 손실 함수임이 명확하다. 이는 dyn_wmse2가 단순히 특정 비율에서 최적의 성능을 보이는 것이 아니라, 결측 환경이 악화되는 상황에서도 예측 성능을 방어하는 구조적 특성을 갖고 있음을 의미한다.

3.5.3 블록 결측 시나리오

본 절에서는 결측이 특정 시간 구간에 연속적으로 발생하는 블록 결측 시나리오에서 각 손실 함수의 성능을 비교·분석한다. 블록 결측은 센서 다운타임, 통신 장애, 장비 점검과 같이 실제 기상 관측 환경에서 빈번히 발생하는 구조적 결측 형태를 모사하며, 랜덤 결측에 비해 시계열 연속성과 정보 손실의 영향이 훨씬 크게 나타난다.

앞선 실험 결과를 바탕으로, 블록 결측 시나리오에서 수행된 성능 비교 실험 결과를 종합적으로 분석한다. 블록 결측은 특정 시간 구간에 결측이 연속적으로 발생하는 구조적 결측 형태로, 랜덤 결측에 비해 시계열 연속성이 크게 훼손되며 입력 정보의 손실이 누적되는 특성을 가진다.

블록 결측은 특정 시간 구간에 결측이 연속적으로 발생하는 구조적 결측 형태로, 랜덤 결측에 비해 시계열 연속성이 크게 훼손되며 입력 정보의 손실이 누적되는 특성을 가진다. 이러한 환경은 실제 기상 관측 시스템에서 발생하는 센서 다운타임, 통신 장애, 장비 점검 상황을 직접적으로 모사하므로, 손실 함수의 강건성과 예측 안정성을 평가하는 것보다 엄격한 조건을 제공한다.

앞선 3.5.3절의 Table 7의 정량적 비교 결과에서 확인한 바와 같이, ANN과 LSTM 두 모델 계열 모두에서 dyn_wmse2

(mask+SHAP)는 모든 결측 비율 구간에서 일관되게 MSE 기준 최상위를 유지하였다. 이는 특정 결측 비율에서의 우수한 성능에 국한된 결과가 아니라, 결측 비율이 점진적으로 증가하는 상황에서도 상대적인 성능 열화를 지속적으로 억제했음을 의미한다.

Fig. 4의 ANN 블록 결측 시나리오 MSE 변화(좌)와 LSTM 블록 결측 시나리오 MSE 변화(우)는 결측 비율 증가에 따른 절대 MSE 변화를 시각적으로 보여준다. 두 그래프에서 공통적으로 관찰되는 특징은, dyn_wmse2(mask+SHAP)의 성능 곡선이 결측 비율 증가에 따라 가장 완만한 기울기로 상승한다는 점이다. 이는 결측 비율이 0.0에서 0.4로 증가하는 동안, 절대적인 성능 저하 폭이 MSE, MAE, Huber와 같은 정적 손실 함수 및 WMSE 계열 손실 함수에 비해 현저히 작음을 의미한다. 이는 결측 비율이 0.0에서 0.4로 증가하는 동안, 절대적인 성능 저하 폭이 MSE, MAE, Huber와 같은 정적 손실 함수 및 WMSE 계열 손실 함수에 비해 현저히 작음을 의미한다. 다시 말해, 블록 결측이 심화될수록 dyn_wmse2는 예측 오차 증가를 가장 효과적으로 억제하는 손실 함수로 작동한다.

반면, 정적 손실 함수들은 결측 비율 증가에 따라 비교적 선형적이거나 일부 구간에서 급격한 성능 저하를 보이며, 결측이 누적되는 환경에서 입력 정보의 신뢰도 변화를 충분히 반영하지 못하는 한계를 드러낸다. 이에 비해 dyn_wmse2는 단순히 오차 크기만을 최소화하는 방식이 아니라, 결측 마스크와 SHAP 기반 중요도를 결합하여 학습 과정에서 신뢰도가 낮은 구간과 예측에 중요한 feature를 구분하고, 이를 손실 계산에 동적으로 반영함으로써 보다 안정적인 성능 변화를 보인다.

또한 WMSE(mask) 및 WMSE(mask×SHAP) 계열 손실 함수는 블록 결측 환경에서 전반적으로 불안정한 거동

Table 7. Block missing performance ANN

MSE	Ratio				
	0.0	0.1	0.2	0.3	0.4
loss					
dyn_wmse2 (mask+SHAP)	0.0400	0.0652	0.0914	0.1160	0.1419
mae	0.0520	0.0750	0.1130	0.1180	0.1600
dyn_wmse (mask+grdinput)	0.0540	0.0850	0.0940	0.1200	0.1530
mse	0.0510	0.0880	0.0940	0.1310	0.1520
huber	0.0550	0.0860	0.1050	0.1320	0.1440
wmse(mask)	0.3140	0.2810	0.3490	0.3420	0.3020
wmse(mask*shap)	0.3020	0.2920	0.3240	0.2980	0.3290

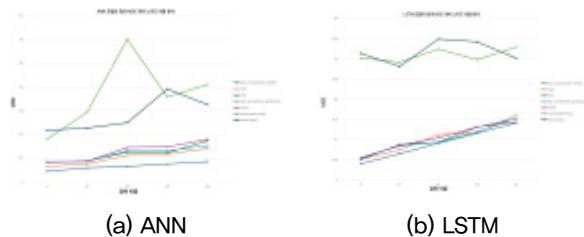


Fig. 4. Performance degradation comparison under increasing block-missing durations

을 나타낸다. 특히 WMSE(mask×SHAP)의 경우, 중요도를 정적으로 적용한 결과 결측이 집중된 구간에서 가중치가 과도하게 증폭되거나 왜곡되는 현상이 관찰되며, 이는 오히려 성능 열화를 가속시키는 방향으로 작용한다. 이러한 결과는 SHAP과 같은 중요도 정보가 동적으로 갱신되지 않고 고정적으로 사용될 경우, 블록 결측과 같은 구조적 결측 환경에서 손실 설계 측면의 한계를 가질 수 있음을 시사한다.

grad×input 기반 dyn_wmse(mask+grad×input) 역시 dyn_wmse2에 비해 상대적으로 가파른 성능 저하 곡선을 보인다. 이는 'gradient 기반 중요도가 국소 민감도(local sensitivity)에 의존하여, 장기 결측으로 인한 입력 분포가 왜곡된 상황에서는 중요도 추정 안정성이 저하되기 쉽다'는 특성을 반영한 결과로 해석할 수 있다.

종합하면, 블록 결측 시나리오에서의 실험 결과는 dyn_wmse2(mask+SHAP)가 결측 비율 증가에 따른 성능 하락을 가장 완만하게 유지하며, 구조적·연속적 결측 환경에서도 ANN과 LSTM 모델 전반에 걸쳐 일관된 성능 우위를 제공함을 명확히 보여준다. 이는 본 연구에서 제안한 동적 손실 함수가 단순한 가중 오차 함수가 아니라, 결측 정보와 feature 중요도를 학습 과정 내부에서 지속적으로 재해석하고 반영하는 설계임을 실험적으로 뒷받침하는 결과라 할 수 있다.

4. 결론

본 연구에서는 기상 시계열 데이터에서 빈번하게 발생하는 결측 문제를 전처리 단계에서 단순히 제거하거나 보정하는 방식이 아닌, 학습 과정 내부에서 결측 특성과 입력 중요도를 동적으로 반영하는 손실 함수 설계 관점에서 접근하였다. 특히 실제 기상 관측 환경에서 관찰되는 결측이 무작위적 누락뿐만 아니라, 센서 다운타임이나 통신 장애로 인해 특정 시간 구간에 연속적으로 발생하는 구조적 결측 형태를 띠는 점에 주목하였다.

이를 위해 전주 및 장수 관측소의 장기간 시계열 분석을 통해 결측 패턴을 정량적으로 분석하고, 결측을 일시적 누락과 장기적 누락으로 구분할 수 있음을 확인하였다. 이러

한 분석 결과를 바탕으로, 결측을 단순한 데이터 오류가 아닌 정보 신뢰도의 변화로 해석하고, 이를 학습 과정에 직접 반영하기 위한 동적 손실 함수 설계를 제안하였다.

제안한 dyn_wmse2(mask+SHAP)는 결측 마스크 기반 신뢰도 가중과 SHAP 기반 특성 중요도를 결합하여, 학습 진행 상황과 입력 특성 변화에 따라 손실 가중치를 동적으로 조정하는 구조를 가진다. 이는 고정 가중치를 사용하는 기존 MSE, MAE, Huber 손실이나 정적 WMSE 계열 손실 함수가 시계열 환경에서 변화하는 입력 중요도와 결측 구조를 충분히 반영하지 못한다는 한계를 보완하기 위한 설계이다.

실험 결과, 랜덤 결측과 블록 결측 시나리오 모두에서 dyn_wmse2는 ANN과 LSTM 두 모델 계열 전반에 걸쳐 모든 결측 비율 구간에서 MSE 기준 최상위 성능을 일관되게 유지하였다. 특히 결측 비율이 증가함에 따라 불가피하게 발생하는 성능 저하 과정에서, dyn_wmse2는 비교 손실 함수 대비 가장 완만한 성능 열화 곡선을 보이며 높은 예측 안정성을 유지하였다. 이는 단순히 평균 성능이 우수한 것을 넘어, 결측 환경 변화에 대한 방어 능력 측면에서 유의미한 차별성을 가진 결과로 해석할 수 있다.

반면, grad×input 기반 동적 가중 손실은 국소 민감도에 의존하는 특성으로 인해 장기 결측 환경에서는 안정적인 중요도 추정에 한계를 보였으며, 정적 WMSE 계열 손실은 블록 결측 환경에서 가중치 왜곡으로 인한 성능 저하가 두드러졌다. 이러한 결과는 중요도 정보를 손실 함수에 반영할 때, 동적 갱신 메커니즘과 안정화 설계가 필수적임을 시사한다.

종합하면, 본 연구는 시계열 예측 문제에서 결측 처리의 초점을 전처리 기법의 고도화가 아닌, 손실 함수 설계와 학습 과정의 구조적 개선으로 전환할 수 있음을 실험적으로 보여준다. 특히 SHAP 기반 중요도를 활용한 동적 손실 함수는 결측 환경이 복잡하고 비정상적인 실제 관측 데이터에서 예측 안정성을 확보하기 위한 효과적인 접근임을 확인하였다.

향후 연구로는 SHAP 계산 비용을 줄이기 위한 근사 기법 도입, 공간적 상호작용을 고려한 그래프 기반 중요도 확장, 그리고 실제 발전량 예측이나 에너지 관리 시스템(EMS) 적용을 통한 실증 검증이 필요하다. 이러한 확장은 본 연구

에서 제안한 동적 손실 함수 프레임워크의 실용성과 일반성을 더욱 강화할 수 있을 것으로 기대된다.

감사의 글

본 과제(결과물)은 2025년도 교육부 및 전북특별자치도의 재원으로 전북RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다(2025-RISE-13-JVU).

References

- [1] Crespo-Turrado, C., Martín-Martín, I., Gutiérrez-Esteban, P., and Rodrigo-Moreno, M., 2014, “Missing data imputation of solar radiation data under different atmospheric conditions”, *Sensors*, **14**(11), 20382-20399.
- [2] International Organization for Standardization, 2018, “ISO 9060:2018 – Solar energy — Specification and classification of instruments for measuring hemispherical solar and direct solar radiation”, Geneva, Switzerland.
- [3] Liu, Q., Mu, Y., Zhao, J., Feng, J., and Wang, B., 2020, “Characterizing packet loss in city-scale LoRaWAN deployment: Analysis and implications”, *Proc. IFIP Networking Conference*, 704-712.
- [4] Wirsinna, A., 2022, “LoRaWAN as part of a smart city strategy”, *Int. J. Innov. Technol. Econ.*, **3**(39), 116-120.
- [5] Choi, C., Jung, H., and Cho, J., 2021, “An ensemble method for missing data of environmental sensor considering univariate and multivariate characteristics”, *Sensors*, **21**(22), 7595, <https://doi.org/10.3390/s21227595>.
- [6] Alejo-Sanchez, L.E., Márquez-Grajales, A., Salas-Martínez, F., Franco-Arcega, A., López-Morales, V., Acevedo-Sandoval, O.A., and Villegas Vega, R., 2025, “Missing data imputation of climate time series: A review”, *MethodsX*, **15**, 103455, <https://doi.org/10.1016/j.mex.2025.103455>.
- [7] Cao, W., Wang, D., Li, J., Zhou, H., Li, L., and Li, Y., 2018, “BRITS: Bidirectional Recurrent Imputation for Time Series”, *Advances in Neural Information Processing Systems (NeurIPS 2018)*, **31**.
- [8] Luo, Y., Cai, X., Zhang, Y., Xu, J., and Yuan, X., 2018, “Multivariate Time Series Imputation with Generative Adversarial Networks”, *Advances in Neural Information Processing Systems (NeurIPS 2018)*, **31**.
- [9] Luo, Y., Zhang, Y., Cai, X., and Yuan, X., 2019, “E²GAN: End-to-End Generative Adversarial Network for Multivariate Time Series Imputation”, *Proc. Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, <https://www.ijcai.org/proceedings/2019/0429.pdf>.
- [10] Oh, E., Kim, T., Ji, Y., and Khyalia, S., 2021, “STING: Self-Attention Based Time-Series Imputation Networks Using GAN”, *Proc. 2021 IEEE International Conference on Data Mining (ICDM)*, 1264-1269.
- [11] Cini, A., Marisca, I., and Alippi, C., 2022, “Filling the Gaps: Multivariate Time Series Imputation by Graph Neural Networks”, *Proc. International Conference on Learning Representations (ICLR)*, <https://openreview.net/pdf?id=kOu3-S3wJ7>.
- [12] Tashiro, Y., Song, J., Song, Y., and Ermon, S., 2021, “CSDI: Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation”, *Proc. Advances in Neural Information Processing Systems (NeurIPS 2021)*, **34**.
- [13] Zhang, H., Fang, L., Wu, Q., and Yu, P.S., 2025, “DIFFPUTER: Empowering Diffusion Models for Missing Data Imputation”, *Proc. International Conference on Learning Representations (ICLR)*, <https://openreview.net/pdf?id=3f11SENSYO>.
- [14] Bishop, G., Si, T., Luebbert, I., Al-Hammadi, N., and Gong, H., 2025, “tBN-CSDI: A Time-Varying Blue Noise-Based Diffusion Model for Time-Series Imputation”, *Bioinform. Adv.*, **5**(1), vbaf225, <https://doi.org/10.1093/bioadv/vbaf225>.
- [15] Simethy, G., Ortiz-Arroyo, D., and Durdevic, P., 2025, “STDiff: A State Transition Diffusion Framework for Time Series Imputation in Industrial Systems”, *Expert Syst. Appl.*, **302**, 130592, <https://doi.org/10.1016/j.eswa.2025.130592>.